Método de los Mínimos Cuadrados Recursivos

Teodoro Alamo[a∗]

Ingeniería de Control
Tercer Curso GITI
Escuela Superior de Ingenieros. Sevilla.

1 Formulación del problema

Supongamos que los escalares y_i, $i = 0, 1, \ldots$ representan la salida (muestreada) de un sistema dinámico. Asumamos asimismo que el valor de la salida y_i puede ser aproximado por una expresión del tipo

$$y_i \approx m_i \theta, \quad i = 0, 1, \ldots,$$

donde $m_i \in \mathbb{R}^{1 \times n_m}$ es un vector fila denominado regresor que está constituido por valores pasados de las entradas y salidas del sistema. Por otro lado, θ es un vector columna donde se concentran los distintos parámetros que determinan la relación entre las entradas y salidas del sistema. Dentro de las entrada podemos asumir las acciones de control y las perturbaciones medibles.

Como ejemplo, supóngase un sistema con salida y y una entrada u y una relación entre la salida y entrada dada por la siguiente ecuación en diferencias:

$$y_i = a_1 y_{i-1} + b_1 u_{i-1} + b_2 u_{i-2} + v_i.$$

El escalar v_i aglutina los errores en la medida, errores de modelado, etc. En este caso, el regresor para un determinado instante de muestreo i es

$$m_i = \begin{bmatrix} y_{i-1} & u_{i-1} & u_{i-2} \end{bmatrix}.$$

El término v_i no se introduce en el regresor puesto que es un término de error que no es conocido a priori. Por otro lado, el vector paramétrico θ sería

$$\theta = \begin{bmatrix} a_1 \\ b_1 \\ b_2 \end{bmatrix}.$$

[∗a] Departamento de Ingeniería de Sistemas y Automática, Universidad de Sevilla, Escuela Superior de Ingenieros, Camino de los Descubrimientos s/n, 41092 Sevilla. Spain. e-mail: `talamo@us.es`

El problema de identificación paramétrica consiste en estimar el vector paramétrico θ de la observación de las salidas y_i y sus respectivos regresores m_i. Nótese que no siempre es posible construir el regresor. Por ejemplo, si se disponen sólo de datos a partir de $i = 0$ en adelante, no se podría construir el regresor m_0 puesto que éste depende de los valores pasados y_{-1}, u_{-1} y u_{-2}. El primer regresor que podría ser construido es m_2, que vendría dado por $\begin{bmatrix} y_1 & u_1 & u_0 \end{bmatrix}$. Por este motivo, es usual que se asuma que el primer par salida regresor disponible es el (y_n, m_n), donde n suele ser mayor cero.

El problema de identificación es especialmente interesante en el contexto de sistemas variantes con el tiempo. En este caso los parámetros que definen la relación entrada-salida cambian de una forma más o menos suave con el tiempo. En este contexto se denota como θ_k la estimación del vector paramétrico en el instante de muestreo k. En los mínimos cuadrados se obtiene θ_k de forma que se minimize de forma ponderada el error cuadrático. Es decir, se minimiza el funcional

$$J_k(\theta) = \sum_{i=n}^{k} \lambda^{k-i}(y_i - m_i\theta)^2.$$

El escalar $\lambda \in (0, 1]$ se denomina factor de olvido. Si se hace igual a la unidad entonces todos los errores cometidos afectan de igual forma. En caso de que λ sea estrictamente menor que uno, los errores recientes (muestras próximas a k) tienen más relevancia que los lejanos en el tiempo. De esta forma se permite que la estimación de los parámetros pueda capturar las variaciones temporales de los mismos. La adecuada elección del valor de λ es crítica para la obtención de resultados satisfactorios. Si definimos

$$Y_k = \begin{bmatrix} y_n \\ y_{n+1} \\ \vdots \\ y_k \end{bmatrix}, \; M_k = \begin{bmatrix} m_n \\ m_{n+1} \\ \vdots \\ m_k \end{bmatrix}, \; W_k = \begin{bmatrix} \lambda^{k-n} & & & & \\ & \lambda^{k-n-1} & & & \\ & & \ddots & & \\ & & & \lambda & \\ & & & & 1 \end{bmatrix},$$

Se obtiene que el índice a minimizar viene dado por

$$J_k(\theta) = (Y_k - M_k\theta)^\top W_k(Y_k - M_k\theta).$$

Denotemos θ_k el valor que minimiza el funcional $J_k(\theta)$. Para obtener el valor óptimo de θ_k analizamos el funcional para una perturbación $\theta_k + \Delta\theta$ respecto del óptimo:

$$\begin{aligned} J_k(\theta_k + \Delta\theta) &= (Y_k - M_k(\theta_k + \Delta\theta))^\top W_k(Y_k - M_k(\theta_k + \Delta\theta)) \\ &= (Y_k - M_k\theta_k)^\top W_k(Y_k - M_k\theta_k) + \Delta^\top\theta M_k^\top W_k M_k\Delta\theta_k \\ &\quad -(Y_k - M_k\theta_k)^\top W_k M_k\Delta\theta - \Delta\theta^\top M_k^\top W_k(Y_k - M_k\theta_k) \\ &= J_k(\theta_k) + \Delta\theta^\top M_k^\top W_k M_k\Delta\theta_k - 2\Delta\theta^\top M_k^\top W_k(Y_k - M_k\theta_k). \end{aligned}$$

Supongamos ahora que elegimos θ_k tal que se eliminen los términos lineales respecto de $\Delta\theta$ (esto es equivalente a hacer cero el gradiente respecto de θ_k):

$$M_k^\top W_k(Y_k - M_k\theta_k) = 0$$

De aquí inferimos que $M_k^\top W_k Y_k = M_k^\top W_k M_k \theta_k$ o equivalentemente:

$$\theta_k = (M_k^\top W_k M_k)^{-1} M_k^\top W_k Y_k. \tag{1}$$

Con esta elección resulta que

$$J_k(\theta_k + \Delta\theta) = J_k(\theta_k) + \Delta\theta M_k^\top W_k M_k \Delta\theta_k \geq J_k(\theta_k),\ \forall \Delta\theta.$$

Esta última ecuación garantiza que la elección $\theta_k = (M_k^\top W_k M_k)^{-1} M_k^\top W_k Y_k$ es óptima en el sentido de los mínimos cuadrados ponderados.

2 Formulación Recursiva

En esta sección se mostrará cómo obtener el valor de θ_k en el instante de muestreo k del valor del vector paramétrico θ_{k-1} obtenido en el periodo de muestreo anterior. Definamos, dado k, la matriz

$$P_k = \left(\sum_{i=n}^{k} \lambda^{k-i} m_i^\top m_i \right)^{-1}. \tag{2}$$

Esta matriz juega un papel muy relevante no sólo en la obtención de θ_k sino también en la caracterización probabilística de las estimaciones paramétricas obtenidas. Nótese que P_k es una matriz simétrica puesto que está definida como la inversa de una matriz simétrica. Como se muestra a continuación, es posible obtener P_k de forma recursiva utilizando para ello P_{k-1}. La siguiente expresión proporciona el valor de P_{k-1} en la muestra $k-1$:

$$P_{k-1} = \left(\sum_{i=n}^{k-1} \lambda^{k-1-i} m_i^\top m_i \right)^{-1}.$$

Por tanto,

$$\begin{aligned}
P_k^{-1} &= \sum_{i=n}^{k} \lambda^{k-i} m_i^\top m_i \\
&= m_k^\top m_k + \sum_{i=1}^{k-1} \lambda^{k-i} m_i^\top m_i \\
&= m_k^\top m_k + \lambda \sum_{i=1}^{k-1} \lambda^{k-1-i} m_i^\top m_i \\
&= m_k^\top m_k + \lambda P_{k-1}^{-1}.
\end{aligned}$$

De esta forma hemos inferido la relación que relaciona P_k con P_{k-1}:

$$P_k = \left(m_k^\top m_k + \lambda P_{k-1}^{-1}\right)^{-1}. \tag{3}$$

El siguiente lema permite obtener una relación algebraica entre P_k y P_{k-1} que evita realizar la inversa de una matriz. El lema que se presenta aquí es un caso particular de un resultado más genérico denominado en la literatura especializada "lema de inversión".

Lema 1 *Supongamos que la matriz simétrica P es invertible. Entonces, dado el escalar $\lambda > 0$ y el vector fila m se tiene que:*

$$\left(m^\top m + \lambda P^{-1}\right)^{-1} = \frac{1}{\lambda}\left(P - \frac{(mP)^\top(mP)}{\lambda + mPm^\top}\right).$$

Demostración:

De la definición de inversa de una matriz inferimos que para demostrar el lema basta con probar la siguiente igualdad:

$$\left(m^\top m + \lambda P^{-1}\right)\left(\frac{1}{\lambda}\left(P - \frac{(mP)^\top(mP)}{\lambda + mPm^\top}\right)\right) = \mathrm{I}.$$

En efecto,

$$\begin{aligned}
\Gamma &= \left(m^\top m + \lambda P^{-1}\right)\left(\frac{1}{\lambda}\left(P - \frac{(mP)^\top(mP)}{\lambda + mPm^\top}\right)\right)\\
&= \frac{1}{\lambda}m^\top(mP) + \mathrm{I} - \frac{m^\top(mPm^\top)(mP)}{\lambda(\lambda + mPm^\top)} - \frac{\lambda m^\top(mP)}{\lambda(\lambda + mPm^\top)}\\
&= \mathrm{I} + \frac{1}{\lambda}m^\top(mP)\left(1 - \frac{mPm^\top}{\lambda + mPm^\top} - \frac{\lambda}{\lambda + mPm^\top}\right)\\
&= \mathrm{I} + \frac{1}{\lambda}m^\top(mP)\left(1 - \frac{\lambda + mPm^\top}{\lambda + mPm^\top}\right)\\
&= \mathrm{I}.
\end{aligned}$$

□

De la aplicación directa del lema anterior a la ecuación (3) obtenemos que

$$P_k = \frac{1}{\lambda}\left(P_{k-1} - \frac{(m_k P_{k-1})^\top(m_k P_{k-1})}{\lambda + m_k P_{k-1} m_k^\top}\right). \tag{4}$$

Esta igualdad tiene muchas implicaciones prácticas. Entre ellas destacamos que la matriz P_k se obtiene del regresor m_k y del valor de P_{k-1} sin la necesidad de realizar la inversa de ninguna matriz. Nótese que esto es algo significativo puesto que P_k se definió en la ecuación (2) como la inversa de una matriz.

A continuación demostramos que P_k^{-1} es igual a la matriz $M_k^\top W_k M_k$, la cual juega un papel central en el cómputo de θ_k.

$$
\begin{aligned}
M_k^\top W_k M_k &= \begin{bmatrix} m_n^\top & \dots & m_{k-1}^\top & m_k^\top \end{bmatrix} \begin{bmatrix} \lambda^{k-n} & & & \\ & \ddots & & \\ & & \lambda & \\ & & & 1 \end{bmatrix} \begin{bmatrix} m_n \\ \vdots \\ m_{k-1} \\ m_k \end{bmatrix} \\
&= \begin{bmatrix} m_n^\top & \dots & m_{k-1}^\top & m_k^\top \end{bmatrix} \begin{bmatrix} \lambda^{k-n} m_n \\ \vdots \\ \lambda m_{k-1} \\ m_k \end{bmatrix} \\
&= \sum_{i=n}^{k} \lambda^{k-i} m_i^\top m_i = P_k^{-1}.
\end{aligned}
$$

Por lo tanto, de la ecuación (1) obtenemos

$$
\begin{aligned}
\theta_k &= (M_k^\top W_k M_k)^{-1} M_k^\top W_k Y_k \\
&= P_k M_k^\top W_k Y_k \\
&= P_k \begin{bmatrix} m_n^\top & \dots & m_{k-1}^\top & m_k^\top \end{bmatrix} \begin{bmatrix} \lambda^{k-n} & & & \\ & \ddots & & \\ & & \lambda & \\ & & & 1 \end{bmatrix} \begin{bmatrix} y_n \\ \vdots \\ y_{k-1} \\ y_k \end{bmatrix} \\
&= P_k \left(\sum_{i=n}^{k} \lambda^{k-i} m_i^\top y_i \right).
\end{aligned}
$$

De la misma manera, el valor de θ_{k-1} viene dado por:

$$\theta_{k-1} = P_{k-1} \left(\sum_{i=n}^{k-1} \lambda^{k-1-i} m_i^\top y_i \right). \tag{5}$$

La formulación recursiva consiste en obtener una expresión que nos permita obtener θ_k del valor de θ_{k-1}. Esto es lo que conseguimos con el siguiente desarrollo.

$$
\begin{aligned}
\theta_k &= P_k \left(\sum_{i=n}^{k} \lambda^{k-i} m_i^\top y_i \right) \\
&= P_k \left(m_k^\top y_k + \sum_{i=n}^{k-1} \lambda^{k-i} m_i^\top y_i \right)
\end{aligned}
$$

$$
\begin{aligned}
&= P_k \left(m_k^\top y_k + \lambda \sum_{i=n}^{k-1} \lambda^{k-1-i} m_i^\top y_i \right) \\
&= \frac{1}{\lambda} \left(P_{k-1} - \frac{(m_k P_{k-1})^\top (m_k P_{k-1})}{\lambda + m_k P_{k-1} m_k^\top} \right) \left(m_k^\top y_k + \lambda \sum_{i=n}^{k-1} \lambda^{k-1-i} m_i^\top y_i \right).
\end{aligned}
$$

Teniendo en cuenta la ecuación (5) obtenemos

$$
\begin{aligned}
\theta_k &= \frac{1}{\lambda} \left(P_{k-1} - \frac{(m_k P_{k-1})^\top (m_k P_{k-1})}{\lambda + m_k P_{k-1} m_k^\top} \right) m_k^\top y_k \\
&\quad + \theta_{k-1} - \frac{(m_k P_{k-1})^\top (m_k \theta_{k-1})}{\lambda + m_k P_{k-1} m_k^\top} \\
&= \frac{1}{\lambda} \left(P_{k-1} m_k^\top - \frac{P_{k-1} m_k^\top (m_k P_{k-1} m_k^\top)}{\lambda + m_k P_{k-1} m_k^\top} \right) y_k \\
&\quad + \theta_{k-1} - \frac{(m_k P_{k-1})^\top (m_k \theta_{k-1})}{\lambda + m_k P_{k-1} m_k^\top} \\
&= \frac{1}{\lambda} P_{k-1} m_k^\top \left(1 - \frac{m_k P_{k-1} m_k^\top}{\lambda + m_k P_{k-1} m_k^\top} \right) y_k \\
&\quad + \theta_{k-1} - \frac{(m_k P_{k-1})^\top (m_k \theta_{k-1})}{\lambda + m_k P_{k-1} m_k^\top} \\
&= \frac{P_{k-1} m_k^\top y_k}{\lambda + m_k P_{k-1} m_k^\top} + \theta_{k-1} - \frac{P_{k-1} m_k^\top (m_k \theta_{k-1})}{\lambda + m_k P_{k-1} m_k^\top} \\
&= \theta_{k-1} + \frac{P_{k-1} m_k^\top (y_k - m_k \theta_{k-1})}{\lambda + m_k P_{k-1} m_k^\top}.
\end{aligned}
$$

Teniendo en cuenta que $y_k - m_k \theta_{k-1}$ es el error cometido en la estimación de y_k con θ_{k-1}, resulta que el valor actualizado de θ_k es igual al valor anterior θ_{k-1} más un término correctivo consistente en el producto del error de estimación $y_k - m_k \theta_{k-1}$ por una ganancia de estimación K_k que adopta la forma:

$$
K_k = \frac{P_{k-1} m_k^\top}{\lambda + m_k P_{k-1} m_k^\top}.
$$

Es decir,

$$
\theta_k = \theta_{k-1} + K_k (y_k - m_k \theta_{k-1}).
$$

Por otra parte, de la expresión de P_k dada por la ecuación (4) tenemos que P_k se puede expresar en términos de K_k y P_{k-1}:

$$
P_k = \frac{1}{\lambda} (\mathrm{I} - K_k m_k) P_{k-1}.
$$

Como resumen de todo lo anterior tenemos que las expresiones que permiten obtener de forma recursiva el valor de θ_k son

$$
\begin{aligned}
K_k &= \frac{P_{k-1} m_k^\top}{\lambda + m_k P_{k-1} m_k^\top} \\
\theta_k &= \theta_{k-1} + K_k (y_k - m_k \theta_{k-1}) \\
P_k &= \frac{1}{\lambda} (\mathrm{I} - K_k m_k) P_{k-1}.
\end{aligned}
$$